

WO 03/059945

PCT/GB02/05941

1

1 Soluble Recombinant Protein Production

2

3 The present invention relates to methods of
4 producing proteins, in particular to methods
5 suitable for high-throughput production of soluble
6 proteins.

7

8 This application describes a methodology for the
9 rapid production of soluble recombinant protein
10 using high-throughput techniques. This method
11 allows the cloning, expression and identification of
12 soluble protein from a given target gene product by
13 a rapid robust method. This ability to produce and
14 analyse soluble recombinant protein in a rapid time
15 period represents a significant advance in an area
16 which has long been considered a significant
17 production bottleneck in the field.

18

19 Introduction

20

21 The recombinant production of protein in bacteria,
22 yeast, insect and mammalian cell lines has become a

1 cornerstone of biological research and the
2 biotechnology industry. Classical biochemical and
3 chromatographical purification techniques usually
4 produce inadequate amounts of a target protein to
5 study its roles or actions. Even if enough of the
6 protein can be purified, it usually involves
7 cumbersome amounts of starting material or tissue
8 and many processing steps are taken before
9 reasonable purification can be achieved.

10

11 Recombinant expression of the target protein
12 bypasses a lot of these problems. By introducing
13 the target protein's gene template to a cell line or
14 bacterial culture, induced overexpression can result
15 in significant levels of that protein being
16 produced. Large amounts of protein make the
17 purification a lot simpler, but the addition or
18 fusion of purification domains or tags allows for a
19 relatively simple one-step purification using
20 affinity chromatography resins.

21

22 Bacteria, and more specifically, *E.coli* are ideal
23 expression vehicles for the production of
24 recombinant protein, as large amounts of foreign
25 protein can be expressed in small culture volumes at
26 low cost in comparison with other methods, for
27 example mammalian cell culture. However, the use of
28 bacteria as expression hosts are not without
29 problems. One of the most troublesome shortcomings
30 of the use of *E.coli* is the production of the
31 recombinant protein in an insoluble form, especially
32 a problem when the target gene is non-bacterial.

1 Generally, insolubility is the result of the
2 production of protein that is not recognised by the
3 folding enzymes, or chaperones, present in the
4 bacterial cytoplasm. The unfolded or misfolded
5 protein will attempt to decrease its own entropy to
6 a minimum, and it is thought that in an effort to
7 hide or mask its hydrophobic residues from the
8 aqueous environment, the protein molecules
9 aggregate. These aggregates are insoluble and are
10 called inclusion bodies. While in the form of
11 inclusion bodies, the protein will have no
12 biological activity and will be impossible to purify
13 using affinity fusion tags. These inclusion bodies
14 can be re-solubilised in chaotropic buffers such as
15 8M urea or 6M guanidine hydrochloride, but then must
16 be slowly dialysed against physiological buffers in
17 an effort to refold and regain biological function.
18 Due to the individual characteristics of each
19 protein, this is a slow and painstaking process that
20 may never produce active or useful protein.
21 Therefore, the ability to quickly produce and screen
22 soluble protein in bacteria such as *E.coli*
23 represents a major step forward in protein
24 biochemistry.

25

26 Summary of the Invention

27

28 The following methodology presented describes a
29 high-throughput process for the cloning, expression
30 and analysis of recombinant soluble protein and
31 protein domains. This process incorporates
32 evaluation and comparison of many factors and

1 conditions known to influence protein solubility at
2 each step in order to guarantee generation of
3 soluble recombinant protein.

4

5 According to the present invention there is provided
6 a method of producing a soluble bioactive domain of
7 a protein the method comprising the step of
8 selecting suitable soluble subunits of a protein and
9 assessing the produced protein for desired activity.

10

11 The method may comprise the steps of amplifying DNA
12 encoding at least one candidate soluble domain,
13 cloning the amplified DNA into at least one
14 expression vector, using each of said vectors into
15 which the DNA has been cloned to each transfect or
16 transform one or more host cell strains, expressing
17 said DNA in one or more host cell strains, and
18 analysing expression products from said host cells
19 for solubility.

20

21 Typically the method comprises the steps of analysis
22 of DNA coding for the protein of interest to
23 identify antigenic soluble domains, designing
24 oligonucleotide primers to amplify DNA encoding the
25 domain, amplifying DNA, cloning the DNA, optionally
26 screening clones for correct orientation of DNA,
27 expressing DNA in expression strains, analysing
28 expression products for solubility, analysing
29 products and production of soluble bioactive protein
30 domain.

31

1 The method optionally comprises the step of
2 producing a soluble bioactive protein domain of said
3 protein of interest.

4
5 In preferred embodiments of the method according of
6 the invention at least three candidate soluble
7 domains are selected and used in the method in
8 parallel. Thus, in preferred embodiments, each stage
9 of the method of the invention is performed for each
10 domain in parallel i.e. primers are designed for
11 each domain in parallel, prior to amplification and
12 ligation of inserts for each insert being performed
13 in parallel prior to propagation of clones being
14 performed in parallel. However, according to this
15 embodiment, although preferred, it is not essential
16 that each stage of the method is completed for all
17 domains prior to the next stage of the method being
18 initiated for one or more domains. There may be
19 slight staggering of stages of the method between
20 domains by e.g. one or two days.

21
22 To further increase the success of the method DNA
23 encoding each selected domain is preferably
24 amplified under at least two, preferably at least
25 three different PCR programs in parallel.

26
27 Preferably, in the method of the invention, the
28 amplified DNA encoding each domain is cloned into a
29 plurality of different expression vectors. Such
30 vectors may include any one or more of a vector
31 capable of encoding a fusion protein with a poly-
32 Histidine tag, a vector capable of conferring tight

1 regulation of translation to impose stringent
2 expression conditions, a vector capable of encoding
3 a fusion protein with a solubility enhancing tag.
4 Typically, the solubility enhancing tag is chosen
5 from the group consisting of a glutathione-S-
6 transferase tag, a dihydrofolate reductase tag, a
7 NusA tag and a SNUT tag.

8

9 In preferred embodiments, the vectors are each
10 transfected or transformed into a plurality of
11 different host cell strains, preferably different *E.*
12 *coli* strains.

13

14 As described below, in developing the method of the
15 present invention, the inventors have developed a
16 novel purification tag based on the gene product of
17 a sortase gene, in particular the *srtA* gene of
18 *Staphylococcus aureus*. This tag, known as SNUT
19 [Solubility eNhancing Unique Tag] has been found to
20 have exceptional activity, enabling the efficient
21 purification of soluble domains of a number of
22 proteins hitherto not able to be isolated
23 efficiently using conventional purification tags.
24 Throughout this specification, reference to a SNUT
25 Tag should be understood to mean a tag derived from
26 a sortase gene product.

27

28 In preferred embodiments, the sortase gene product
29 is a gene product of the *srtA* gene of *Staphylococcus*
30 *aureus*.

31

1 Accordingly, in preferred embodiments of the method
2 of the invention, vectors capable of encoding a
3 fusion protein with a SNUT tag are used.

4

5 However, utility of the SNUT Tag is not limited to
6 use in the method of the present invention. Indeed
7 in a second independent aspect of the invention,
8 there is provided a purification tag comprising a
9 sortase, e.g srtA, gene product.

10

11 Also provided is the use of a sortase, e.g srtA,
12 gene product as a purification tag.

13

14 Furthermore, according to a third aspect of the
15 invention, there is provided an expression construct
16 for the production of recombinant polypeptides,
17 which construct comprises an expression cassette
18 consisting of the following elements that are
19 operably linked: a) a promoter; b) the coding region
20 of a DNA encoding a sortase, eg srtA gene product as
21 a purification tag sequence; c) a cloning site for
22 receiving the coding region for the recombinant
23 polypeptide to be produced; and d) transcription
24 termination signals.

25

26 According to a fourth aspect of the invention, there
27 is provided a method for producing a polypeptide,
28 comprising: a) preparing an expression vector for
29 the polypeptide to be produced by cloning the coding
30 sequence for the polypeptide into the cloning site
31 of an expression construct according to the third
32 aspect of the invention; b) transforming a suitable

1 host cell with the expression construct thus
2 obtained; and c) culturing the host cell under
3 conditions allowing expression of a fusion
4 polypeptide consisting of the amino acid sequence of
5 the purification tag with the amino acid sequence of
6 the polypeptide to be expressed covalently linked
7 thereto; and, optionally, d) isolating the fusion
8 polypeptide from the host cell or the culture medium
9 by means of binding the fusion polypeptide present
10 therein through the amino acid sequence of the
11 purification tag.

12

13 The expression construct, herein referred to as
14 pSNUT, may be made by modification of any suitable
15 vector to include the coding region of a DNA
16 encoding a sortase. In preferred embodiments, the
17 expression construct is based on the pQE30 plasmid.

18

19 A sample of pSNUT was deposited with the National
20 Collections of Industrial and Marine Bacteria Ltd.
21 (NCIMB), 23 St Machar Drive, Aberdeen, Scotland AB24
22 3RY on 23 December 2002 under accession no NCIMB
23 41153.

24

25 In a fifth aspect, there is provided a fusion
26 polypeptide obtained by the method of the fourth
27 aspect of the invention.

28

29 In preferred embodiments, the sortase, e.g.
30 srtA, gene product (SNUT) is encoded by the
31 nucleotide sequence shown in Figure 8 or a variant
32 or fragment thereof. Preferably, the srtA gene

1 product comprises amino acids 26 to 171 of the SrtA
2 sequence shown in Figure 8 or a variant or fragment
3 thereof.

4

5 Variants and fragments for use in the invention
6 preferably retain the functional capability of the
7 polypeptide i.e. ability to be used as a
8 purification tag. Such variants and fragments which
9 retain the function of the natural polypeptides, can
10 be prepared according to methods for altering
11 polypeptide sequence known to one of ordinary skill
12 in the art such as are found in references which
13 compile such methods, e.g. Molecular Cloning: A
14 Laboratory Manual, J. Sambrook, et al., eds., Second
15 Edition, Cold Spring Harbor Laboratory Press, Cold
16 Spring Harbor, New York, 1989, or Current Protocols
17 in Molecular Biology, F. M. Ausubel, et al., eds.,
18 John Wiley & Sons, Inc., New York.

19

20 A variant nucleic acid molecule shares homology
21 with, or is identical to, all or part of the coding
22 sequence discussed above. Generally, variants may
23 encode, or be used to isolate or amplify nucleic
24 acids which encode, polypeptides which are capable
25 of ability to be used as a purification tag.

26

27 Preferred variants include one or more of the
28 following changes(using the annotation of AF162687):
29 nucleotide 604 AAG causing an amino acid mutation of
30 KAR; nucleotide 647 AAG, codon remains K, therefore
31 a silent mutation; nucleotide 966 GAA causing an
32 amino acid mutation of GAQ.

1
2 Variants of the present invention can be artificial
3 nucleic acids (i. e. containing sequences which have
4 not originated naturally) which can be prepared by
5 the skilled person in the light of the present
6 disclosure. Alternatively they may be novel,
7 naturally occurring, nucleic acids, which may be
8 isolatable using the sequences of the present
9 invention. Thus a variant may be a distinctive part
10 or fragment (however produced) corresponding to a
11 portion of the sequence provided in Figure 8. The
12 fragments may encode particular functional parts of
13 the polypeptide.
14
15 The fragments may have utility in probing for, or
16 amplifying, the sequence provided or closely related
17 ones.
18
19 Sequence variants which occur naturally may include
20 alleles or other homologues (which may include
21 polymorphisms or mutations at one or more bases).
22 Artificial variants (derivatives) may be prepared by
23 those skilled in the art, for instance by site
24 directed or random mutagenesis, or by direct
25 synthesis. Preferably the variant nucleic acid is
26 generated either directly or indirectly (e. g. via
27 one or amplification or replication steps) from an
28 original nucleic acid having all or part of the
29 sequences of Figure 8. Preferably it encodes a
30 polypeptide which can be used as a purification
31 tag.
32

1 The term 'variant' nucleic acid as used herein
2 encompasses all of these possibilities. When used in
3 the context of polypeptides or proteins it indicates
4 the encoded expression product of the variant
5 nucleic acid.

6
7 Homology (i. e. similarity or identity) may be as
8 defined using sequence comparisons are made using
9 FASTA and FASTP (see Pearson & Lipman, 1988. Methods
10 in Enzymology 183 : 6398). Parameters are preferably
11 set, using the default matrix, as follows :

12 Gapopen (penalty for the first residue in a gap) :-
13 12 for proteins/-16 for DNA

14 Gapext (penalty for additional residues in a gap) :-
15 2 for proteins/-4 for DNA

16 KTUP word length : 2 for proteins/6 for DNA.

17 Homology may be at the nucleotide sequence and/or
18 encoded amino acid sequence level. Preferably, the
19 nucleic acid and/or amino acid sequence shares at
20 least about 60%, or 70%, or 80% homology, most
21 preferably at least about 90%, 95%, 96%, 97%, 98% or
22 99% homology with the sequence shown in Figure 8.

23
24 Thus a variant polypeptide in accordance with the
25 present invention may include within the sequence
26 shown in Figure 8, a single amino acid or 2, 3, 4,
27 5, 6, 7, 8, or 9 changes, about 10, 15, 20, 30, 40
28 or 50 changes. In addition to one or more changes
29 within the amino acid sequence shown, a variant
30 polypeptide may include additional amino acids at
31 the C terminus. and/or N-terminus.

32

1 Naturally, regarding nucleic acid variants, changes
2 to the nucleic acid which make no difference to the
3 encoded polypeptide (i. e. 'degeneratively
4 equivalent') are included within the scope of the
5 present invention.

6
7 Changes to a sequence, to produce a derivative, may
8 be by one or more of addition, insertion, deletion
9 or substitution of one or more nucleotides in the
10 nucleic acid, leading to the addition, insertion,
11 deletion or substitution of one or more amino acids
12 in the encoded polypeptide. Changes may be by way of
13 conservative variation, i. e. substitution of one
14 hydrophobic residue such as isoleucine, valine,
15 leucine or methionine for another, or the
16 substitution of one polar residue for another, such
17 as arginine for lysine, glutamic for aspartic acid,
18 or glutamine for asparagine. As is well known to
19 those skilled in the art, altering the primary
20 structure of a polypeptide by a conservative
21 substitution may not significantly alter the
22 activity of that peptide because the side-chain of
23 the amino acid which is inserted into the sequence
24 may be able to form similar bonds and contacts as
25 the side chain of the amino acid which has been
26 substituted out. This is so even when the
27 substitution is in a region which is critical in
28 determining the peptides conformation.

29
30 Also included are variants having non-conservative
31 substitutions. As is well known to those skilled in
32 the art, substitutions to regions of a peptide which

1 are not critical in determining its conformation may
2 not greatly affect its activity because they do not
3 greatly alter the peptide's three dimensional
4 structure.

5
6 In regions which are critical in determining the
7 peptides conformation or activity such changes may
8 confer advantageous properties on the polypeptide.
9 Indeed, changes such as those described above may
10 confer slightly advantageous properties on the
11 peptide e. g. altered stability or specificity.

12
13 The invention is exemplified with reference to the
14 following non limiting description and the
15 accompanying figures in which

16
17 Figure 1 illustrates the basic protocol used in an
18 embodiment of the invention.

19
20 Figure 2 shows a putative timetable for the process
21 from analysis of the protein to expression of
22 immunisation-ready protein.

23
24 Figure 3 shows selected domains for amplification
25 from *in silico* analysis. Representation of a
26 candidate protein for the expression platform, in
27 this case Jak1 (human). Four fragments have been
28 chosen by analysis as depicted.

29
30 Figure 4 shows amplification of target domains of
31 the human gene *SOCS6* by PCR. Agarose electrophoresis
32 results of the amplification of three fragments from

1 a cDNA clone of the human gene SOCS6. (a) shows
2 domain a (lane 1); domain b (lane 2) and domain c
3 (lane 3) results of amplification using the
4 anticipated annealing temperature as calculated by
5 primer design software as described. Lanes 4-6 show
6 the same amplification procedures using 5% DMSO for
7 inserts a, b and c respectively. (b).
8 Amplification of domains a, b and c using touchdown
9 program in the absence of DMSO (1, 2 and 3) and in
10 the presence of 5% DMSO (lanes 4, 5 and 6). (c).
11 Amplification of same domains using 50 °C annealing
12 temperature, again in the absence of DMSO (1, 2 and
13 3), and in the presence of 5% DMSO (lanes 4, 5 and
14 6).
15
16 Figure 5 shows denaturing dot-blot analysis of
17 expression clones of fragments of MAR1 in pQE30.
18
19 Figure 6 shows SDS-PAGE and Western blot analysis of
20 soluble lysates. Total protein staining of a 4-20%
21 Bio-Rad Criterion SDS-PAGE gel using chloroform (a),
22 followed by subsequent western blotting of same gel
23 and detection of bands using monoclonal antibody-HRP
24 to poly-histidine tag (b). Results correspond to
25 individual clones expressing NusA-Yotiao protein
26 fusions.
27
28 Figure 7 shows a ribbon Diagram of *Staphylococcus*
29 *aureus* sortase. Ribbon diagram of the putative
30 structure of *S. aureus* SrtA protein (minus its N-
31 terminal membrane anchor). SNUT represents the
32 portion of this structure between the two yellow

1 arrows as shown. The yellow ball signifies a Ca^{2+}
2 ion, essential for the biological activity of this
3 protein. This diagram is taken from Ilangoan et
4 al., 2001, PNAS 98 (11) 6056
5 (doi:10.1073/pnas.101064198)

6
7 Figure 8 shows the Nucleotide Sequence and amino
8 acid sequence of SNUT fragment

9
10 (a) This is the determined sequence of SNUT. The
11 fragment was cloned into pQE30 using the BamHI site
12 of this vector. When in the wanted orientation,
13 insertion results in the inactivation of the
14 upstream cloning site, therefore allowing any
15 subsequent cloning of target inserts with the
16 downstream BamHI site (see (b) for restriction map
17 of sequence).

18
19 Figure 9 illustrates qualitative purification
20 results using the SNUT fusion tag. (a) shows the
21 elution profile on SDS-PAGE of SNUT-Jak1 using AKTA
22 Prime native histag purification. Successful
23 elution of SNUT-Jak1 construct is signified by the
24 white arrow. (b) shows the elution profile on SDS-
25 PAGE of SNUT-MAR1 using AKTA Prime native histag
26 purification. Successful elution is shown by the
27 arrow. (c) shows the same gel stained in (b)
28 western blotted and detected using poly-histidine-
29 HRP antibody. This is confirmation that the eluted
30 species in (b) is actually SNUT-MAR1, of expected
31 molecular weight.

32

1 Template analysis and primer design

2

3 The high throughput process begins with the analysis
4 of the DNA coding for the protein of interest.

5 Software packages such as Vector NTI (Informax, USA)

6 and BLASTP(<http://www.ncbi.nlm.nih.gov/BLAST/>), p-

7 fam (www.sanger.ac.uk/pfam) and TM pred

8 (www.hgmp.mrc.ac.uk) may be used to identify

9 complete domains within the protein that

10 significantly increase the likelihood of

11 antigenicity and/or solubility when expressed as a

12 subunit of the original protein coding sequence. In

13 order to increase the possibility of identifying a

14 soluble domain, preferably multiple sub-domains,

15 more preferably at least three sub-domains, for

16 example 3 to 9 sub-domains are identified for

17 processing. This has proven optimal to produce

18 soluble protein with the majority of proteins

19 expressed using the method of the invention.

20

21 The next step in the process is to design

22 oligonucleotide primers to amplify the selected sub-

23 domains. Primer design may be aided by use of

24 commercially available software packages such as the

25 internet software package Primer3 ([http://www-](http://www-genome.wi.mit.edu/genome)

26 [genome.wi.mit.edu/genome](http://www-genome.wi.mit.edu/genome)

27 [software/other/primer3.html](http://www-genome.wi.mit.edu/genome/software/other/primer3.html)) (Whitehead Institute

28 for Biomedical Research), Vector NTI

29 (www.informaxinc.com) and DNASIS (Hitachi Software

30 Engineering Company) (www.oligo.net). These packages

31 allow full control over all aspects of primer

32 design, ranging from primer length, homology to

1 optimal annealing temperature of the PCR reaction
2 itself.

3

4 Typically primers for use in the method of the
5 invention are in the range 10-50 base pairs in
6 length, preferably 15 to 30, for example 20 base
7 pairs in length, with annealing temperatures in the
8 range 45-72°C, for example 50-60°C, more
9 conveniently 55-60°C. Primers may be synthesised
10 using standard techniques or may be sourced from
11 commercial suppliers such as Invitrogen Life
12 Technologies (Scotland) or MWG-Biotech AG (Germany).

13

14 PCR of Insert

15

16 The desired inserts which encode the selected sub-
17 domains are amplified using the primers designed
18 specifically for that target gene using standard PCR
19 techniques. The template DNA for amplification can
20 be in the form of plasmid DNA, cDNA or genomic DNA,
21 depending on whatever is appropriate or indeed
22 available. Any suitable DNA polymerase may be used,
23 for example, Platinum Taq, Pfu (www.stratagene.com)
24 or Pfx (www.invitrogen.com). . Any suitable PCR
25 system may be used. In the examples detailed
26 herein, the Expand High Fidelity PCR system (Roche,
27 Basel, Switzerland), was used with working stocks of
28 each primer made (10pMol/ μ l).

29

30 In preferred embodiments of the invention, several
31 different thermocycler conditions are used with each
32 set of primers. This increases the chance of the PCR

1 working without having to individually optimise each
2 new primer set. Typically the following three
3 programs are used in the method of the invention:

- 4
- 5 1. A standard PCR programme using the recommended
6 annealing temperature provided with the
7 primers.
 - 8 2. A standard PCR programme using 50°C as the
9 temperature for annealing.
 - 10 3. A touchdown PCR programme, where the annealing
11 temperature starts at a high temperature e.g
12 65°C for 10 cycles and then gradually decreases
13 the annealing temperature to 50°C over the
14 subsequent e.g 15 cycles.

15
16 Buffer conditions may be adjusted as required, for
17 example with respect to magnesium ion concentration
18 or addition of DMSO for the amplification of
19 difficult templates.

20
21 The PCR products are then visualised using standard
22 techniques, for example on a 1.5% agarose gel
23 stained with Ethidium Bromide and the bands are cut
24 out of the gel and purified using Mini elute gel
25 extraction Kit (Qiagen, Crawley, England).

26
27 **Expression Vectors**

28
29 Amplified DNA inserts are subsequently cloned into
30 expression vectors using techniques dictated by the
31 multiple cloning sites of the vector in question.

1 Such techniques are readily available to the skilled
2 person.

3

4 In order to maximise the successful generation of
5 soluble antigen, the amplified DNA coding for each
6 target protein domain is preferably cloned into a
7 plurality of different expression vectors. This
8 allows the generation of a library of novel
9 expression constructs which can then simultaneously
10 be screened for the high level production of soluble
11 protein. Each construct will have different
12 properties due to attachment of 'tag' domains, which
13 are designed to increase expression and solubility.

14

15 Any suitable expression system can be used in the
16 method of the invention. Preferably, the expression
17 system is prokaryotic. Preferably at least two
18 expression vectors, preferably three, most
19 preferably 4 to 5 vectors are used for each of the
20 constructs in the method of the invention.
21 Preferably, vector combinations are chosen to allow
22 the same cloning methodologies to be used
23 simultaneously as this allows a much more rapid
24 entry in expression trials.

25

26 Suitable vectors for use in the method of the
27 invention include one or more of the following:

28

29 I. Vectors that will generate fusion protein with a
30 poly-Histidine tag (his-tag, hexahistidine tag, or
31 his-patch). The expressed His tag can be situated
32 at either the N or C terminus of the protein, or

1 even internally. Examples include the pQE series
2 from Qiagen, Valencia, CA; pET 14-19, Novagen,
3 Madison, WI. A poly-histidine tag is an non-natural
4 amino acid sequence with unusual and specific
5 chelation properties with metal bivalent ions such
6 as Ni^{2+} and Cu^{2+} . Immobilised metal affinity
7 chromatography (IMAC) exploits this property to
8 allow the specific purification of proteins
9 containing this tag, therefore making it an
10 extremely useful purification tool.

11

12 II. Vectors that confer tight regulation of
13 translation to impose stringent expression
14 conditions especially for proteins that are toxic to
15 a prokaryotic host. An example of such a vector is
16 the pQE80 vector, Qiagen. Tight regulation is
17 absolutely essential for the production of some
18 proteins, especially proteins foreign to the
19 bacterial host which are more likely to have toxic
20 effects to the bacterial host. Some high-level
21 expression systems are not particularly stringent
22 and leaky expression may occur without induction,
23 causing bacterial hosts to be killed before a
24 culture has reached a great enough density to
25 sustain expression of a toxic gene.

26

27 III. Vectors that will generate fusion proteins with
28 a solubility enhancing tag such as glutathione-S-
29 transferase (examples include the pGEX series,
30 Amersham Biosciences, Uppsala, Sweden; pET41/2,
31 Novagen) or NusA (pET43, Novagen). These tags have
32 been identified as proteins of a highly soluble

1 nature in E. coli and confer their soluble
2 characteristics to proteins attached to them as
3 fusion partners.
4

5 IV. Vectors that encode fusion partners that
6 facilitate the expression of small or poorly
7 expressed proteins including glutathione-S-
8 transferase and dihydrofolate reductase (Amersham
9 Biosciences and Qiagen respectively). Some
10 proteins, due to the composition of the coding DNA
11 are only poorly expressed in bacteria. In some cases
12 they may not be produced at all. Tags such as GST
13 and DHFR can aid such expression if incorporated as
14 N-terminal fusions to help generate adequate amounts
15 of a target protein, where no protein would be
16 expressed if the template was only the target DNA.
17

18 V. Vectors that encode SNUT. [Solubility eNhancing
19 Unique Tag], for example pSNUT. This tag is based on
20 the sequence of a trans-peptidase found on the
21 surface of gram-positive bacteria. This protein is
22 highly soluble, and expressed as very high levels.
23 As described below, the inventors have found that
24 SNUT is an ideal fusion tag for conferring
25 solubility and expression levels to target protein
26 fragments. SNUT may be cloned into any suitable
27 vector. For the purposes of the results shown in
28 this application, the sequence incorporating the
29 SNUT fragment is cloned into pQE30 in a manner
30 allowing full use of the multiple cloning site (MCS)
31 of this vector for downstream gene insertions.
32

1 Development of pSNUT

2
3 Occasionally, due to the varying nature of proteins,
4 the production of soluble protein has remained
5 elusive. In fact in some cases, production of
6 protein can be a problem due to differences in the
7 machinery of bacterial cells. During the
8 development of this high-throughput expression
9 platform, the need for a more versatile tag than is
10 available currently on the market became evident.

11
12 The inventors found that a tag based on the *srtA*
13 gene product from *Staphylococcus aureus* is highly
14 soluble nature, reacts well to purification schemes
15 and expresses particularly well. It was
16 hypothesised that the incorporation of a portion or
17 domain of this protein could represent a useful
18 fusion tag in the present method, and indeed the
19 expression of any poorly soluble protein in *E. coli*.
20 Using NMR studies, the 3D structure of this protein
21 has been predicted and is shown in Figure 7. We
22 hypothesised that by taking a portion of this
23 structure, we could make a manipulatable protein
24 tag, but not disturb its tertiary structure enough
25 to reduce its highly favourable characteristics
26 listed above. The region of this protein used as a
27 solubility-enhancing tag is depicted by two arrows.

28
29 To make this tag compatible with the other vectors
30 and systems being used on the platform, this SNUT
31 tag was cloned into pQE30 as described earlier.
32 However, it may be cloned into any suitable

1 expression vector. Positive clones may be identified
2 by denaturing dot blots, SDS-PAGE and Western
3 blotting. Final confirmation of these clones was
4 provided by DNA sequencing, and the sequence of the
5 multiple cloning region of the resultant vector is
6 shown in Figure 8.

7
8 Variances in the sequence of the SNUT domain were
9 observed from the sequence for SrtA that has been
10 logged in Genbank (AF162687). The variances are
11 (using the annotation of AF162687) nucleotide 604
12 AAG causing an amino acid mutation of KAR;
13 nucleotide 647 AAG, codon remains K, therefore a
14 silent mutation; nucleotide 966 GAA causing an amino
15 acid mutation of GAQ.

16
17 Preliminary trials and native purification showed
18 that the SNUT fragment was very soluble and its
19 characteristics were in no way diminished by
20 truncation, thus showing that SNUT could represent a
21 useful tag domain (data not shown). As described in
22 the Examples, to fully test the abilities of SNUT,
23 we then chose two proteins were soluble protein
24 production had proved impossible using conventional
25 methods and using the other expression systems of
26 the method of the present invention. Surprisingly,
27 we found that, using pSNUT in the method of the
28 invention, these proteins could be produced in
29 soluble form.

30

1 Accordingly, in preferred embodiments of the method
2 of the invention, at least one of the vectors
3 encodes SNUT.

4

5 Clone Propagation

6

7 Target insert/expression vector ligations are
8 propagated using standard transformation techniques
9 including the use of chemically competent cells or
10 electro-competent cells. The choice of the host
11 cell and strain for transformation is dependent on
12 the characteristics of the expression vectors being
13 utilised.

14

15 In the method of the invention, bacterial cells,
16 for example, *Escherchia coli*, are the preferred host
17 cells. However, any suitable host cell may be used.

18 In preferred embodiments, the host cells are

19 *Escherchia coli*.

20

21 In preferred embodiments of the present invention,
22 in order to further maximise the chances of success
23 in isolating a soluble protein, one or more,
24 preferably all of the vectors are used to each
25 transfect or transform a plurality of different host
26 cell strains. The set of host cell strains for
27 individual vector may be the same or different from
28 the set used with other vectors.

29

30 In a particularly preferred embodiment of the
31 invention, each vector is transformed into three *E.*
32 *coli* strains (for example, selected from

1 Rosetta(DE3)pLacI, Tuner(DE3)pLacI, Origami BL21
2 (DE3)pLacI and TOP10F, Qiagen).

3

4 Where the vectors are pQE based vectors, TOP10F'
5 cells are preferred for the propagation and
6 expression trials of such vectors. The present
7 inventors have identified this strain as a more
8 superior strain for these vectors than either of the
9 recommended strains by the supplier (M15(pREP4) and
10 SG13009(pREP4)), in terms of ease of use and culture
11 maintenance (only one antibiotic required as to two
12 with M15(pREP4) or SG13009(pREP4) (www.qiagen.com)).
13 Other F' strains such as XL1 Blue can be used, but
14 are inferior to the TOP10F' strain, due to lack of
15 expression regulation (results not shown). The use
16 of TOP10F' (Invitrogen) for the propagation and/or
17 expression pQE based vectors forms an independent
18 aspect of the present invention. Other F' strains
19 such as XL1 Blue may also be used, but are inferior
20 to the TOP10F'.

21

22 After transformation, cells are plated out onto
23 selection plates and propagated for the development
24 of single colonies using standard conditions.

25

26 Propagation of Cells

27

1 In preferred embodiments, the colonies are used to
2 inoculate wells in a 96 well plate.

3

4 Routinely, 6-48 clones for each insert-vector
5 ligation are taken and propagated in culture micro-
6 titre plates containing up to 500 µl of media.

7

8

9

10 Typically, each well may contain 200 µl of LB broth
11 with the appropriate antibiotics. Each plate is
12 dedicated to one strain of E. coli or other host
13 cell which alleviates the problems of different
14 growth rates. The necessary controls are also
15 included on each plate. The plates are then grown
16 up, preferably at 37°C or any other temperature as
17 appropriate to the particular host cell and vector,
18 with shaking, until stationary phase is reached.
19 This is the primary plate.

20

21 From the primary plate a secondary plate is seeded
22 and then grown to log phase. Typically, the
23 secondary plate is seeded using 'hedgehog'
24 replicators. Determination of positive clones from
25 these plates may be undertaken using functional
26 studies. According to the conditions and reagents
27 required, protein production is then induced, and
28 cultures propagated further. Most vectors are under
29 the control of a promoter such as T7, T7lac or T5,
30 and can be easily induced with IPTG during log phase
31 growth. Typically, cultures are propagated in a
32 peptone-based media such as LB or 2YT supplemented

1 with the relevant antibiotic selection marker.
2 These cultures are grown at temperatures ranging
3 from 4-40 °C, but more frequently in the range of
4 20-37 °C depending on the nature of the expressed
5 protein, with or without shaking and induced when
6 appropriate with the inducing agent (usually log or
7 early stationary phase). After induction, growth
8 propagation can be continued for 1-16 hours for a
9 detectable amount of protein to be produced.

10

11 The primary plate is preferably stored at 4°C as a
12 reference, until the process is complete.

13

14 Colony Screening for Inserts in Correct Orientation

15

16 The method of the invention may include the step of
17 testing transformants for correct orientation of the
18 inserts.

19

20 Although all colony selecting and picking can be
21 done manually, automated colony pickers are
22 preferred. Automated colony pickers such as the
23 BioRobotics BioPick allow for the uniform and
24 reproducible selection of clones from transformation
25 plates. Clone selection determinants can be set to
26 ensure picking colonies of a standardised size and
27 shape. After picking and plate inoculation,
28 propagation of clones can be carried out as
29 described above.

30

31 Identification of positive clones can be achieved
32 through a variety of methods, including standard

1 techniques such as digestion analysis of plasmid
2 DNA; colony PCR and DNA sequencing. Alternatively,
3 in a preferred embodiment, the novel method of dot-
4 blotting described herein for the identification of
5 positive clones may be used in place of such
6 traditional techniques, prior to final confirmation
7 by DNA sequencing. The use of this method in the
8 platform presented here is not essential in the use
9 of this platform over existing screening
10 methodologies, but represents a rapid, reproducible
11 and robust detection method. The protocol described
12 here is a new protocol for an existing method for
13 which commercially available equipment (Bio-Rad
14 DotBlot) can be purchased.

15
16 This particular method is useful for the rapid
17 detection or presence of recombinant protein and
18 allows for a determination of all clones
19 irrespective of solubility and conformation. This
20 is useful at this stage, because conformational
21 structures can inhibit the detection of tag domains
22 if they are not presented properly on the surface of
23 the protein. This can occur as easily with both
24 soluble and insoluble protein.

25
26 For example, after growth on the micro-titre plates
27 is complete, the plate is centrifuged at 4000 rpm
28 for 10 minutes at 4°C to harvest the bacterial
29 cells. The supernatant is removed and the cell
30 pellets are re-suspended in 50 µl lysis buffer (10
31 mM Tris.HCl, pH 9.0, 1mM EDTA, 6 mM MgCl₂)
32 containing benzonase (1 µl/ml). The plate is

1 subsequently incubated at 4°C with shaking for 30
2 minutes. A sample (10 µl) of the cell lysate is
3 added to 100 µl buffer (8 M urea, 500 mM NaCl, 20 mM
4 sodium phosphate, pH 8.0) and incubated at room
5 temperature for 20 minutes. Samples are then
6 applied to a BioDot apparatus (BioRad) containing
7 nitrocellulose membrane (0.45µM pore size) in
8 accordance with the manufacturers' instructions.
9 The membrane is removed and transferred into
10 blocking reagent (3% w/v; Bovine serum albumin in
11 TBS) for 30 minutes at room temperature. The blot
12 is washed briefly with TBS then incubated in a
13 primary antibody, specific to the tag being used for
14 the subset of expression clones. Depending on the
15 nature of the primary i.e., whether or not it has a
16 horse radish peroxidase (HRP) reporter function,
17 will depend on whether the use of a secondary is
18 required. For detection of specific binding the
19 membrane is then washed 2x 5 minutes in TBS followed
20 by 1x 5 minute wash in 10 mM Tris.HCl pH7.6.
21 Detection of specifically bound antibody is
22 disclosed by the addition of chromogenic substrate
23 (6 mg diaminobenzidine in 10 ml 10 mM Tris.HCl pH
24 7.6 containing 50 µl 6% H₂O₂) . The reaction is
25 stopped by thorough rinsing in water. Positive
26 clones identified by this procedure can then be
27 confirmed by DNA sequencing of the expression
28 construct using now industry-standard techniques and
29 equipment such as ABI and Amersham Biosciences.

30

31 Sequencing

32

1 The sequencing reactions may be performed using
2 techniques common in the art using any suitable
3 apparatus. For example, sequencing may be performed
4 on the cloned inserts, using the Big Dye Terminator
5 cycle sequencing kits (Applied Biosystems,
6 Warrington, UK) and the specific sequencing primer
7 run on a Peltier Thermal cycler model PTC225 (MJ
8 Research Cambridge, Mass). The reactions may be run
9 on Applied Biosystems - Hitachi 3310 Sequencer
10 according to the manufacturer's instructions. These
11 sequences are checked to ensure that no PCR
12 generated errors have occurred.

13

14 **Assessment of Solubility of Positive Clones**

15

16 The cells of the positive clones may then be
17 harvested and soluble and insoluble protein
18 detected.

19

20 Any suitable techniques known in the art can be used
21 to separate soluble and insoluble protein, such as
22 the use of centrifugation, magnetic bead
23 technologies and vacuum manifold filtrations.
24 Typically, however, the separated proteins are
25 ultimately analysed by acrylamide gel and western
26 blotting. This confirms the presence of recombinant
27 protein at the correct size.

28

29 In one embodiment, contents of each well in the 96
30 well plate are transferred into a Millipore 0.65 μm
31 multi-screen plate. The plate is placed on a vacuum
32 manifold and a vacuum is applied. This draws off

1 the culture medium to waste. The cells are then
2 washed with PBS (optional), again the vacuum is
3 applied to remove the PBS. The multi-screen plate is
4 removed from the manifold and bacterial cell lysis
5 buffer (containing DNase) (50 μ l) is added to each
6 well. The plate is incubated at room temperature
7 for 30 minutes with shaking to facilitate lysis of
8 the cells. A fresh 96 well microtitre plate is
9 placed inside the vacuum manifold and the multi-
10 screen plate is placed above it. When a vacuum is
11 applied the contents of each well are drawn into the
12 micro-titre plate below. The vacuum only needs to
13 be applied for 20 seconds. The collected lysate
14 contains the soluble fraction of expressed protein.
15 A sample of the collected lysate may subsequently
16 analysed by SDS-PAGE and Western blotting to confirm
17 both the presence and correct molecular weight of
18 the target protein.

19

20 The use of SDS-PAGE and Western blotting can be
21 expensive and time consuming, especially when
22 numerous samples must be analysed for each
23 construct. In light of this we have developed a
24 protocol whereby one gel can be used for both total
25 protein staining and western blotting. This
26 represents a significant improvement in this
27 methodology and obviously allows cost saving, and
28 precise comparisons can be made with regard to total
29 protein and western blotting as both sets of results
30 come from the one gel.

31

1 The basis of this protocol is in the ability to use
2 chloroform and UV light to stain protein on an SDS-
3 PAGE gel (Kazmin et al., Anal Biochem, 2002, 301(1)
4 91-6; doi:10.1006/abio.2001.5488). We have used
5 this technique to great effect as it allows for the
6 extremely rapid staining of a SDS-PAGE gel in less
7 than a tenth of the time taken using other more
8 traditional staining methods such as Commassie
9 Brilliant Blue and Collodial Blue stains. We then
10 decided to take this observation a step further and
11 analyse the ability of a chloroform-stained gel to
12 be used in Western blotting. This would not be
13 expected to work as other stained gels result in the
14 fixing of the protein to the gel and subsequent
15 inability to transfer the protein during blotting.
16 This expectation is coupled to the fact that
17 chloroform is not compatible with western blotting
18 equipment (Bio-Rad SD blotter user's manual).
19 However, fortuitously, we have discovered that with
20 a wash of the chloroform-stained gel in double-
21 distilled water, to remove excess chloroform, and
22 after subsequent soaking in transfer buffer,
23 proteins were effectively transferred during western
24 blotting in contrast to expectations. This transfer
25 was no-less effective than from a gel that has not
26 been pre-stained with chloroform and UV light.
27 Figure 6 primarily shows results relating to the
28 production of soluble protein by the platform, but
29 also shows the ability to use the chloroform-stained
30 SDS-PAGE derived western blot for the identification
31 of proteins, without any apparent damage caused to
32 the proteins.

1

2 Th use of a chloroform-stained SDS-PAGE derived
3 western blot for the identification of proteins
4 forms another aspect of the present invention.

5

6 Scale-Up and Purification

7

8 This analysis provides a picture of the expression
9 status of the clones on each plate. Using this
10 analysis, positive soluble protein expressing clones
11 can be identified for the production of soluble
12 recombinant protein for a given target protein. The
13 clones may be selected and their growth scaled up
14 e.g. to 5 ml scale, using the saved primary plate as
15 an inoculum. Parameters that may be taken into
16 consideration in deciding on the appropriate culture
17 to select for scale-up include the desirability of
18 specific regions for the production of an antigen,
19 the overall expression levels of the clone and
20 factors that may affect affinity purification such
21 as amino acid composition.

22

23 Example 1.

24

25 Overview of Process

26

27 Figure 1 illustrates the basic protocol used in an
28 embodiment of the invention. The DNA coding for the
29 protein of interest is analysed to identify target
30 domains which may enhance solubility. For each
31 insert, multiple primers are designed and used to
32 amplify the chosen nucleotide sequences. For each

1 primer set, the PCR reaction is performed under
2 three different thermocycler conditions: a standard
3 PCR programme using the recommended annealing
4 temperature provided with the primers; a standard
5 PCR programme using 50°C as the temperature for
6 annealing; and a touchdown PCR programme, where the
7 annealing temperature starts at 65°C for 10 cycles
8 and then gradually decreases the annealing
9 temperature to 50°C over the subsequent 15 cycles.

10

11 Example 2 Expression construct design

12

13 Figure 3 is a diagrammatic representation of the
14 protein Jak1. Using pfam, the position of distinct
15 domains was established. Further analysis of these
16 domains was then carried out using Tmpred and the
17 Kyle and Dolittle hydrophobicity algorithm to
18 determine the usefulness of these domains as soluble
19 antigens. From this tentative analysis, four
20 domains were selected for amplification and
21 expression analysis.

22

23 Example 3 Parallel Amplification of DNA Sequences

24 Under Different PCR Conditions Enables Rapid

25 Amplification of Inserts of Interest

26

27 Based on preliminary *in silico* analysis, primers
28 specific for a target protein were designed and used
29 to amplify domains selected for analysis. Figure 4
30 shows the amplification of portions of human SOCS6
31 gene from a cDNA plasmid clone using three programs:

- 1 1. A standard PCR programme using the recommended
- 2 annealing temperature provided with the
- 3 primers.
- 4 2. A standard PCR programme using 50°C as the
- 5 temperature for annealing.
- 6 3. A touchdown PCR programme, where the annealing
- 7 temperature starts at a high temperature e.g
- 8 65°C for 10 cycles and then gradually decreases
- 9 the annealing temperature to 50°C over the
- 10 subsequent e.g 15 cycles.
- 11 a) shows domain a (lane 1); domain b (lane 2) and
- 12 domain c (lane 3) results of amplification using the
- 13 anticipated annealing temperature as calculated by
- 14 primer design software. Lanes 4-6 show the same
- 15 amplification procedures using 5% DMSO for inserts
- 16 a, b and c respectively. (b). Amplification of
- 17 domains a,b and c using touchdown program in the
- 18 absence of DMSO (1,2 and 3) and in the presence of
- 19 5% DMSO (lanes 4,5 and 6). (c). Amplification of
- 20 same domains using 50 °C annealing temperature,
- 21 again in the absence of DMSO (1, 2 and 3), and in
- 22 the presence of 5% DMSO (lanes 4,5 and 6). It is
- 23 clear from these results how much more effective the
- 24 use of varying protocols (4b and 4c) is over the
- 25 basic protocol using the pre-determined annealing
- 26 temperatures. These results show the requirement of
- 27 different programs to guarantee the amplification of
- 28 certain inserts, even with gene specific DNA
- 29 primers, as no strict rules can be applied for the
- 30 amplification of DNA for every different gene
- 31 target.

1 Furthermore, the manipulation of the Mg^{2+} and DMSO in
2 the reaction buffer may be useful for the guaranteed
3 amplification of some gene fragments, as seen in
4 Figure 4. In the present example, no amplification
5 of a cancer antigen DNA was successful without the
6 addition of DMSO, which was added in order to
7 disrupt secondary structure and cause some
8 denaturing. This allows primers to anneal to some
9 difficult templates prior to elongation by the DNA
10 polymerise during PCR.

11
12 These results depict the high-throughput nature of
13 the method of the invention, even at a DNA level.
14 These procedures allow the rapid amplification of
15 all gene inserts

16 17 Example 4 Dot blotting

18
19 The optional use of dot-blotting in the method of
20 the invention has proven to be an invaluable tool
21 for the preliminary evaluation of clones for protein
22 expression. Figure 5 shows the results of a
23 denaturing dot-blot analysis of expression clones of
24 fragments of murine antigen receptor MAR1 in pQE30.
25 using the method of the invention. The blot depicts
26 the expression of all 4 target fragments designed in
27 pQE30, and clearly shows the levels of poly-
28 histidine tagged protein in each well. All detection
29 was achieved using horse radish peroxidase conjugate
30 to a poly-histidine tag monoclonal antibody (Sigma).
31 Rows A and B are 24 individual clones of insert 1 in
32 pQE30. Rows C and D represent insert 2; rows E and

1 F represent insert 3 and G and H represent insert 4.
2 Presence of purple product on an individual dot
3 signifies positive detection of the presence of
4 poly-histidine tag and therefore a positive clone.

5

6 **EXAMPLE 5 Evaluation of Soluble Protein From**
7 **yotiao.**

8

9 In this example, results are shown for the
10 expression and analysis of the mammalian gene
11 yotiao. Gene specific primers were designed and
12 used for the amplification of the target regions and
13 these were then cloned into pQE30, pQE80, pGEX and
14 pET43.1a using the following protocol.

15

16 Vectors (500 ng) were restricted with BamHI (20
17 units) and SalI (20 units) in the presence of calf
18 intestinal alkaline phosphatase (CIP) (2 units), gel
19 purified and quantified using standard methods.
20 Purified PCR fragments (100 ng) were restricted with
21 BamHI (5 units) and SalI 5 units), gel purified,
22 quantified, and then used in a ligation reaction
23 with the restricted vector again using standard T4
24 DNA ligase methods (Ready-to-Go T4 DNA ligase,
25 Amersham Biosciences). A sample of the ligation
26 reaction (1 µl) was then used to transform the
27 appropriate competent bacterial cells (TOP10F' were
28 used here for the pQE vectors, a modification of the
29 manufacturers recommendations; BL21(DE3)pLysE for
30 pET43.1a and TOP10F' for pGEX-Fus). Transformants
31 were selected on LB/ampicillin (100 µg/ml) for the
32 pQE and pGEX-Fus vectors and

1 LB/ampicillin/chlorophenicol/glucose for pET43.1 (50
2 µg/ml, 32 µg/ml and 1% respectively) overnight at
3 28 °C.

4
5 A Cambridge BioRobotics BioPick instrument was used
6 for the picking of 24 colonies from each of the
7 transformant plates into flat-bottomed and lidded
8 micro-titre plates. For this screen there were 3
9 inserts in 4 vectors, resulting in a total of 288
10 clones picked. All pQE30, 80 and pGEX-Fus clones
11 were used to inoculate 150 µl of LB (containing
12 100µg/ml ampicillin) (see Figure 1), and these were
13 allowed to grow overnight at 37 °C. For the
14 pET43.1a clones, LB containing 1% glucose, 50 µg/ml
15 ampicillin and 34 µg/ml chloramphenicol were used
16 for propagation. These pET43.1a clones were grown
17 overnight at 28 °C. From this plate, secondary
18 plates were seeded using 'hedgehog' replicators, and
19 these are again grown up to log phase prior to
20 induction with IPTG and being left to grow
21 overnight.

22
23 A secondary plate was then prepared by the
24 inoculation of 200 µl of LB containing the required
25 supplements with 10 µl of the overnight primary
26 culture. These were then grown at 37 °C (for the
27 pQE30, 80 and pGEX-Fus constructs) and 28 °C (for
28 the pET43.1a clones). Once an optical density (OD)
29 of 0.25 at A550 was reached, IPTG (final
30 concentration, 1 mM) is added to induce expression
31 of the recombinant protein. Culture propagation was

1 continued for another 4 hours prior to harvesting of
2 bacterial cells.

3
4 After clones expressing specific recombinant protein
5 have been identified, the solubility of these
6 proteins has to be established prior to clone
7 selection for purification. This can be performed a
8 number of ways including the use of centrifugation
9 and automation-friendly vacuum manifold separations.
10 The results shown here were obtained using
11 methodologies based around the use of vacuum-
12 assisted filtration to separate soluble and
13 insoluble protein. The filtrates that were produced
14 from the method described were then analysed by SDS-
15 PAGE and Western blotting to confirm the production
16 of a recombinant protein of the correct anticipated
17 molecular weight.

18
19 Figure 6 shows the examination of screened-clone
20 soluble extracts by SDS-PAGE and western blotting.
21 These particular results are for the expressed
22 products of the bacterial gene yotiao from the
23 pET43.1a vector (producing Yotiao fragments as NusA
24 fusion proteins). The SDS-PAGE gel shows the clear
25 presence of expressed soluble protein in the
26 lysates, which is confirmed to contain poly-
27 histidine tags on the accompanying western blot.
28 The results in Figure 6 are proof of the
29 effectiveness of the method presented here. The
30 production of soluble protein using one of the
31 expression systems, pET43.1a is clearly visible,
32 thus allowing identification of clones suitable for

1 scale-up cultures and subsequent purification. The
2 production of soluble Yotiao protein fragments from
3 the other systems was tried (pQE30; pQE40 and
4 pQE80), but proved unsuccessful. Clones expressing
5 soluble Yotiao were identified and then confirmed by
6 DNA sequencing within 3 weeks of receiving the cDNA
7 template for the gene.

8
9 These results collectively show the power and
10 utility of the platform. Normally, expression of
11 such a protein would be carried out in just a basic
12 vector such as pQE30 alone, and inability to produce
13 soluble protein using this system, which is also
14 part of the platform, exemplifies the power of the
15 platform to guarantee soluble recombinant protein
16 production.

17
18 **Example 7 Design and Construction of SNUT Expression**
19 **Tag**

20
21 Based on analysis of the amino acid sequence and
22 predicted structure of SrtA_{ΔN}, it was decided to
23 amplify the region of amino acids 26 to 171 of the
24 SrtA sequence. Amplification was conducted using
25 the forward primer 5' TTTTGTAGATCTAAACCACATATCGAT
26 and the reverse primer 5'
27 TTTTGTGGATCCATCTAGAACTTCTAC. This product was then
28 digested with BglI and BamHI and ligated into pQE30
29 vector which had also been digested with BamHI to
30 form the pSNUT vector. The ligation mix was
31 transformed into TOP10F' cells and single colonies
32 propagated on LB agar containing 100 µg/ml

1 ampicillin. Clones with the *srtA* fragment in the
2 correct orientation were screened by expression
3 analysis and positive clones identified using the
4 denaturing dot-blot assay described earlier.

5
6 The sequence encoding the SNUT tag was cloned into
7 pQE30 as described earlier and positive clones
8 identified by denaturing dot blots, SDS-PAGE and
9 Western blotting. Final confirmation of these
10 clones was provided by DNA sequencing, and the
11 sequence of the multiple cloning region of the
12 resultant vector is shown in Figure 8. Variances in
13 the sequence of the SNUT domain were observed from
14 the sequence for *SrtA* that has been logged in
15 Genbank (AF162687). The variances are (using the
16 annotation of AF162687) nucleotide 604 AAG causing
17 an amino acid mutation of KAR; nucleotide 647 AAG,
18 codon remains K, therefore a silent mutation;
19 nucleotide 966 GAA causing an amino acid mutation of
20 GAQ.

21 22 **Example 8 Trials of SNUT Expression Constructs**

23
24 Target inserts were cloned into the pSNUT vector
25 using primer construction and digestion of resulting
26 PCR amplifications with *Bam*HI and *Sal*I as described
27 earlier. pSNUT was digested with *Bam*HI in a similar
28 manner and the target inserts cloned as described.
29 Clones were screened using the denaturing dot-blot
30 system and then analysed with SDS-PAGE and western
31 blotting. Positive clones were used for preparative
32 200 ml LB cultures containing 100 µg/ml ampicillin

1 and induced as described earlier. This was grown to
2 an optical density of 0.5 at A_{550} at 37 °C.
3 Expression of SNUT was then induced with the
4 addition of IPTG (final concentration, 1 mM) and
5 left to grow for another 4 hours. Cells were then
6 harvested by centrifugation at 5K rpm for 15
7 minutes. Cells were re-suspended in 30 ml PBS
8 containing 0.1% Igepal and lysis induced by two
9 freeze-thaw cycles. The suspension was then
10 sonicated and centrifuged at 5K rpm for 15 minutes.
11 The soluble supernatant was transferred to a fresh
12 container and filtered through a 0.8 μ m disc filter
13 to remove final cell debris. This solution was then
14 applied to a Ni^{2+} charged IMAC column (Amersham
15 Biosciences HiTrap Chelating column, 1 ml) using an
16 AKTA Prime low pressure chromatography system and
17 column was then treated using a standard native his-
18 tag purification protocol involving washing of
19 column with 20 mM sodium dihydrogen phosphate pH 8.0
20 containing 10 mM imidazole, 500 mM NaCl, and elution
21 of soluble his-tagged proteins using 20 mM sodium
22 dihydrogen phosphate pH 8.0 containing 500 mM
23 imidazole, 500 mM NaCl... Elution fractions were
24 then analysed on an SDS-PAGE gel (4-20% SDS-PAGE
25 Bio-Rad Criterion gel), which was stained with
26 chloroform as described earlier. This gel was then
27 subsequently western blotted and the his-tagged
28 protein detected with anti-poly-histidine monoclonal
29 antibody as described earlier.
30
31 Preliminary trials and native purification showed
32 that the SNUT fragment was very soluble and its

1 characteristics were in no way diminished by
2 truncation, thus showing that SNUT could represent a
3 useful tag domain (data not shown). To fully test
4 the abilities of SNUT, we then chose two proteins
5 for which soluble protein production had proved
6 impossible using the other expression systems in
7 which SNUT was not used as a tag. These were murine
8 MAR1 and human Jak1. Clones were prepared and
9 selected using the method as described in the
10 Examples above and positive clones were subsequently
11 grown and induced at 37 °C. These were then treated
12 to identical native histag purifications. Both
13 proteins behaved very favourably under standard
14 purification conditions as can be seen from the
15 purification profiles in Figure 9. For both these
16 trial proteins, this was the first example of such
17 purification under soluble conditions. The
18 production of these proteins using conventional
19 techniques has failed to produce any soluble
20 protein, irrespective of expression system or growth
21 conditions used (data not shown). However, as
22 described in this example, when the protein
23 fragments were expressed in pSNUT, soluble proteins
24 can be surprisingly obtained.

25

26 The effectiveness of SNUT as a fusion protein is
27 even more significant when it is considered that no
28 special growth conditions were required for the
29 generation of soluble protein. This is remarkable
30 when one considers the protein expressionist's
31 standard GST tag which is not even soluble itself
32 when expressed at 37 °C; 28 °C is required before

1 even the generation of GST on its own without any
2 target protein is observed.

3

4 In this application we have demonstrated that our
5 high throughput cloning and expression platform can
6 rapidly identify clones that express soluble
7 protein. This is achieved through the use of a
8 number of expression vectors coupled with a range of
9 target fragments. That coupled with our expression
10 conditions; sample processing and analysis ensure
11 that soluble antigen is generated. As can be seen
12 from the results presented, the production of a
13 soluble mammalian protein in *E. coli* can be
14 troublesome and requires the application of several
15 different methodologies, or expression systems and
16 conditions in order to guarantee a successful
17 outcome. The protocols detailed in this
18 specification are the ideal automation-ready platform
19 for generation of such soluble protein. This
20 platform offers not only the generation of soluble
21 protein, but also in a rapid, reproducible and
22 robust manner.

23

24 All documents referred to in this specification are
25 herein incorporated by reference. Various
26 modifications and variations to the described
27 embodiments of the inventions will be apparent to
28 those skilled in the art without departing from the
29 scope and spirit of the invention. Although the
30 invention has been described in connection with
31 specific preferred embodiments, it should be
32 understood that the invention as claimed should not

1 be unduly limited to such specific embodiments.
2 Indeed, various modifications of the described modes
3 of carrying out the invention which are obvious to
4 those skilled in the art are intended to be covered
5 by the present invention.

6

7